

SWLS 网络测验中可靠性变量的测量不变性检验*

王道阳^{1,2}, 刘争光²

(1. 安徽师范大学教育科学学院, 芜湖 241000; 2. 北京师范大学中国基础教育质量监测协同创新中心, 北京 100875)

摘要: 测量不变性在自我报告问卷或量表的心理测量应用中非常重要, 是跨组比较的前提条件。测量不变性检验模型包括无任何约束的分组验证性因素分析 (M_{group})、形态的不变性 (M1)、负荷的不变性 (M2)、截距的不变性 (M3)、严格不变性 (M4)、因子方差 - 协方差的不变性 (M5) 以及潜均值的不变性 (M6)。以生活满意度量表 (SWLS) 为例, 针对 1343 名大学生 (年龄 17 - 25 岁, 20.01 ± 1.53), 进行有急事需要处理 (否 vs. 是), 答题时感受 (积极情绪 vs. 消极情绪), 噪音水平 (无噪音 vs. 有噪音), 答题用时 (长 vs. 短), 性别 (男 vs. 女), 户口 (非农业户口 vs. 农业户口) 等不同组别的生活满意度量表 (SWLS) 完全因素不变性检验。结果表明: (1) 是否有急事需要处理的不变性成立 ($\Delta\chi^2 = 0.49 \sim 10.59, p > 0.05$); (2) 答题时感受不变性部分成立, M5、M6 模型不变性不成立 ($\Delta\chi^2_{(1)} = 3.96, 20.89, p < 0.05$); (3) 噪音水平不变性部分成立, M3 与 M4 模型不变性检验不成立 ($\Delta\chi^2_{(4)} = 14.75, \Delta\chi^2_{(5)} = 23.91, p < 0.05$); (4) 答题用时不变性不成立 ($\Delta\chi^2 = 11.01 \sim 41.95, p < 0.05$); (5) 性别的不变性部分成立, M4 模型不变性检验不成立 ($\Delta\chi^2_{(5)} = 64.40, p < 0.05$); (6) 户口的不变性部分成立, M6 模型不变性检验不成立 ($\Delta\chi^2_{(1)} = 11.49, p < 0.05$)。

关键词: 测量不变性; 结构不变性; 生活满意度量表

中图分类号: B841.2

文献标识码: A

文章编号: 1003 - 5184(2018)01 - 0080 - 06

1 引言

近些年, 随着计算机网络、移动网络的普及, 传统的纸笔测验逐渐被网络测验或在线测验 (Internet or On-line) 所取代 (Dodou & Winter, 2014; Raffaelli et al., 2016)。这是由时代发展、科技进步所决定的, 也因为网络测验本身相对于纸笔测验有着巨大优势。这些优势表现为: (1) 更加经济, 且增加数据收集效率。比如, 减少纸张印刷, 符合节约环保理念, 也控制了成本。(2) 减少了数据录入、清理过程。通过计算机控制可以大大降低数据出错率, 比如, 被试出生时间 (年、月、日), 可以限定“年”为 4 个字符, 限定“月”为“1 - 12”等, 通过网络计算机实现质量的控制; 同时, 答题结束时数据在后台就同时生成, 减少了传统纸笔测验后再次编码、录入、核查、清理数据的过程。(3) 数据取样更大。通过网络计算机数千个样本的数据与数百个样本数据处理的时间几乎是相当的 (Hauk, 2015)。(4) 可以控制观察者偏差 (Observer bias)、反应偏差 (response bias) 等无关因素 (Davis, 1999)。(5) 答题形式更易于被试接受。互动性增强, 可以通过发红包、抽奖的形

式激励被试参与; 同时, 可以记录被试完成答题的时长; 被试答题的时空选择更加自由。

然而, 自我报告的问卷或量表, 绝大多数都是通过纸笔测验编制而成。能否把通过纸笔测验编制而成的问卷或量表, 直接通过网络测验来应用? 也就是说, 网络测验的结果与纸笔测验是否具有对等性 (Measurement Equivalence / Invariance)。特别是, 网络测验对被试答题时的情境和条件是无法控制的。而自我报告测验在不同情境、条件下, 可能会导致测量特性发生变化, 就会带来测量不变性问题 (Hardre, Crowson, & Xie, 2012)。于是, 近些年有关纸笔测验与网络测验的不变性研究报告有很多, 且大多数研究结论认为纸笔测验与网络测验不是完全对等的 (Cieciuch et al., 2015; Hauk, 2015; Lewis, Watson, & White, 2009; Meade, Lautenschlager, & Michels, 2007)。这些研究, 客观上对于把笔纸版的问卷或量表运用到网络测验中起到了积极的推动作用, 也界定了网络测验对测量工具的使用条件。然而, 这些研究仅仅把纸笔测验与网络测验进行方法上的不变性验证, 往往忽视了纸笔测验与网络测验

* 基金项目: 国家社会科学基金教育学青年课题 (CBA120108)。
通讯作者: 刘争光, E-mail: zhengguang_liu@126.com。

之所以有不对等性背后的因素。因为,纸笔测验相对于网络测验最重要的优势就是对于被试测试情境与条件的控制。既然网络测验已有着巨大优势,如果再能把网络测验情境与条件控制得以提升,那么网络测验就具有纸笔测验无可比拟的优势。因此,研究重点应该是如何在网络测验中控制测验情境、条件。已有研究报告网络测验中被试测验场地噪音影响、测试时的情绪体验、答题时长、答题时是否有急事处理以及答题场所等都是影响测验特性的重要影响因素(Hardre et al., 2012)。基于此,可以对于上述网络测验中的情境、条件变量进行不变性检验。如果不变性不成立,则说明某种条件下问卷或量表的测量特性发生变化,这种条件下的数据需要进行处理或清理。所以,通过对假设影响网络测验的情境变量进行不变性检验就可以发现网络测验可靠性控制的方法。

不变性是自我报告问卷不同组别比较的前提条件。测量不变性是对两组或多组问卷条目(观测变量)之间不变性的检验,而结构不变性是对两组或多组问卷维度(潜变量)之间不变性的检验,两者合在一起统称为完全因素不变性(Factorial Invariance)(Byrne, Shavelson, & Muthén, 1989)。完全因素不变性被Byrne等提出之后,主要基于多组验证性因素分析(Multi-group Confirmatory Factor Analysis, MCFA)理论(Meade, & Lautenschlager, 2004)。MCFA检验的参数模型主要包含:无任何约束的分组验证性因素分析(M_{group})、协方差的不变性(M_0 , Invariant Covariance)、形态的不变性(M_1 , Configural Invariance)、负荷的不变性(M_2 , Metric Invariance)、严格不变性(M_4 , Strict Invariance)、因子方差-协方差不变性(M_5 , Factor Variance/Covariance Invariance)、潜均值的不变性(M_6 , Latent Mean Invariance)(Vandenberg & Lance, 2000)。

主观幸福感是人们对其生活现状、质量所做的情感性和认知性的评价,包含情感和认知两个成份。生活满意度量表(the Satisfaction With Life Scale, SWLS)是针对主观幸福感知成份的测评(Diener et al., 1999)。关于SWLS在性别、年龄等人口学变量的不变性检验研究有很多,但随着网络测验的普及,也有研究者通过网络答题方式开展生活满意度量表调查,并且与纸质答题方式进行了测量不变性验证。结果表明,网络测验和纸笔测验之间存在弱不变性,但只存在部分的强不变性和部分的严格不

变性(蔡华俭等, 2008)。随着网络普及,网络测验具有答题不受时间、空间的束缚、数据处理便捷的优点,越来越受到研究者的重视。与纸质测验有主试对答题环境、答题时心理感受进行控制不同,网络测验时这些都不能适当控制或者使之处于同一水平。因此,在网络测验时,就需要设计一些被试答题时噪音水平、心理感受、是否着急(有其他要紧的事等着要做),以及答题所用时间来了解其他变量的干扰,以便对数据进行清理、比较。那么,就带来当被试在不同噪音水平、心理感受等情境下,不同组的自我报告量表结果是具有可比性的问题。也就是说,首先需要解决这些不同组自我报告量表的测量不变性问题。SWLS测量的生活满意度在某种意义上与主观认知有关,而此次研究进行的不变性检验,主要就是对被试在测试时周围环境的认知的两组不变性检验。所以,测试时当前周围环境的认知会不会对幸福感的主观认知产生影响,会不会造成对周围环境的认知不同组被试之间对生活满意度理解产生差异,这需要通过完全因素不变性检验才能确定。

2 研究方法

2.1 研究对象

运用计算机网络和手机网络(扫描微信二维码)的方式,利用网络平台征集答题调查了1343名大学生,年龄17~25岁(20.01 ± 1.53)。其中男生846名(63.00%),女生497名(37.00%);户口性质为非农业户口355人(26.40%),农业户口988人(73.60%)。

2.2 问卷

背景问卷。包括被试年龄、性别、户口性质等人口学变量;也包括被试对答题时周围环境与心理感受的评估,主要有“你现在有事急需处理吗?如,去上课、老师让我去办公室找他等”“你当前心理感受是什么?”“你的周围是否有噪音?噪音是否影响答题?”等。此外,系统自动记录被试答题时间,将作为后期分析的一个重要变量。

生活满意度量表。该量表主要是测评自己对当前生活的满意程度,选项包括非常不同意到非常同意,共7个等级(1~7);量表有5个题目,“我的生活大致符合我的理想”“我的生活状况非常圆满”“我满意自己的生活”“直到现在为止,我都能够得到我在生活上希望拥有的重要东西”“如果我能重新活过,差不多没有东西我想改变”(Diener et al., 1985)。中国香港研究者,将中文版生活满意度量

表应用到香港大学生中,其内部一致性系数 α 为0.78,分半信度为0.70(Wang, Yuen, & Slaney, 2009)。此次研究采用的生活满意度量表,其内部一致性系数 α 为0.90。

2.3 数据分析

针对生活满意度量表,进行有急事需要处理(否、是)、答题时感受(积极情绪、消极情绪)、噪音水平(无、有)、答题用时(长:均数以上、短:均数以下)等不同组别的完全因素不变性检验。验证模型包括无任何约束的分组验证性因素分析(M_{group})、形态的不变性(M1)、负荷的不变性(M2)、截距的不变性(M3)、严格不变性(M4)、因子方差-协方差的不变性(M5)以及潜均值的不变性(M6)。拟合指数使用 χ^2/df 、CFI、TLI、90% CI、RMSEA、SRMR;模型之间

比较使用 $\Delta\chi^2$,即 χ^2 变化量。 $\Delta\chi^2$ 服从自由度变化量(Δdf)的 χ^2 分布。当 $\Delta\chi^2$ 显著时,两模型不变性检验假设不成立,反之则成立。

3 结果与分析

3.1 不同组别生活满意度量表的描述性统计

不同组别(SWLS总分)以及总样本(SWLS总分及各个题目)均分和标准差,以及峰度和偏度值见表1。结果显示,偏度在0.02~0.47之间,峰度0.14~0.80之间,数据完全符合正态分布的标准(峰度绝对值在0~2、偏度绝对值在0~7,可以认为数据是正态分布)(Curran, West, & Finch, 1996)。总样本的内部一致性 α 系数为0.90,不同组别的内部一致性 α 系数也在0.86~0.92之间。

表1 不同组别生活满意度量表的描述性统计

组别		人数	%	M	SD	偏度	峰度	α
有事急需处理	不是(group1)	975	72.60%	19.26	6.73	0.25	-0.31	0.90
	是的(group2)	368	27.40%	19.65	7.13	0.03	-0.37	0.91
答题时感受	积极情绪(group1)	1032	76.80%	19.79	6.63	0.16	-0.38	0.89
	消极情绪(group2)	311	23.20%	17.99	7.33	0.38	-0.14	0.92
噪音水平	无噪音(group1)	374	27.80%	19.73	7.00	0.34	-0.20	0.91
	有噪音(group2)	969	72.20%	19.23	6.77	0.12	-0.41	0.90
答题用时	时间长(group1)	868	64.60%	20.16	7.12	0.14	-0.40	0.91
	时间短(group2)	475	35.40%	17.93	6.04	0.08	-0.50	0.86
全样本	总分	1343	100.00%	19.37	6.84	0.19	-0.33	0.90
	题目1	—	—	3.95	1.57	-0.02	-0.74	—
	题目2	—	—	3.96	1.55	0.06	-0.70	—
	题目3	—	—	4.15	1.57	-0.11	-0.77	—
	题目4	—	—	4.05	1.60	-0.04	-0.80	—
	题目5	—	—	3.26	1.78	0.47	-0.80	—

注:答题时感受分类,积极情绪包括放松或开心,消极情绪包括生气、沮丧、紧张或烦恼。

3.2 不同组别生活满意度量表完全因素不变性验证结果分析

表2是不同组别生活满意度量表完全因素不变性各模型拟合指数汇总。从不同组别的分组无任何

约束的验证性因素分析的结果来看,均表现出CFI>0.95、TLI>0.95,虽然部分模型RMSEA的值大于0.10,但SRMR<0.08,而RMSEA更易于收到样本数量的影响。因此,总体上看,分组的模型拟合较好。

表2 不同组别生活满意度量表完全因素不变性模型的拟合指数

模型	χ^2	df	$\Delta\chi^2$	Δdf	CFI	ΔCFI	90% CI	RMSEA	SRMR
有事急需处理									
M_{group1}	44.45	5	—	—	0.990	—	0.067-0.115	0.090	0.020
M_{group2}	54.16	5	—	—	0.960	—	0.126-0.204	0.163	0.030
M1	98.61	10	—	—	0.980	—	0.095-0.136	0.115	0.023
M2	100.25	14	1.64	4	0.980	0.000	0.079-0.114	0.096	0.026
M3	102.89	18	2.44	4	0.981	0.001	0.068-0.100	0.084	0.026
M4	113.48	23	10.59	5	0.979	-0.002	0.063-0.091	0.077	0.027
M5	114.84	24	1.36	1	0.979	0.000	0.062-0.089	0.075	0.036
M6	115.33	25	0.49	1	0.979	0.000	0.060-0.087	0.073	0.033

续表 2

模型	χ^2	df	$\Delta\chi^2$	Δdf	CFI	ΔCFI	90% CI	RMSEA	SRMR
答题时感受									
M _{group1}	65.93	5	—	—	0.980	—	0.086 - 0.133	0.109	0.024
M _{group2}	21.99	5	—	—	0.990	—	0.062 - 0.151	0.105	0.020
M1	87.92	10	—	—	0.982	—	0.088 - 0.129	0.108	0.023
M2	93.54	14	5.62	4	0.982	0.000	0.075 - 0.110	0.092	0.029
M3	101.60	18	8.06	4	0.981	-0.001	0.068 - 0.099	0.083	0.032
M4	110.49	23	8.89	5	0.980	-0.001	0.062 - 0.090	0.075	0.030
M5	114.45	24	3.96*	1	0.979	-0.001	0.061 - 0.089	0.075	0.052
M6	135.34	25	20.89**	1	0.974	-0.005	0.068 - 0.095	0.081	0.093
噪音水平									
M _{group1}	22.36	5	—	—	0.990	—	0.058 - 0.139	0.096	0.019
M _{group2}	66.72	5	—	—	0.980	—	0.090 - 0.138	0.113	0.025
M1	89.08	10	—	—	0.982	—	0.088 - 0.130	0.109	0.023
M2	91.45	14	2.37	4	0.982	0.000	0.074 - 0.109	0.091	0.027
M3	106.20	18	14.75**	4	0.980	-0.002	0.070 - 0.101	0.085	0.028
M4	130.11	23	23.91**	5	0.976	-0.004	0.070 - 0.097	0.083	0.033
M5	130.41	24	0.30	1	0.976	0.000	0.068 - 0.095	0.081	0.033
M6	130.83	25	0.42	1	0.976	0.000	0.066 - 0.093	0.079	0.033
答题用时									
M _{group1}	45.83	5	—	—	0.990	—	0.072 - 0.124	0.097	0.018
M _{group2}	40.78	5	—	—	0.970	—	0.090 - 0.159	0.123	0.034
M1	86.61	10	—	—	0.982	—	0.087 - 0.128	0.107	0.025
M2	106.30	14	19.69**	4	0.979	-0.003	0.082 - 0.117	0.099	0.047
M3	137.42	18	31.12**	4	0.972	-0.007	0.084 - 0.115	0.099	0.047
M4	179.37	23	41.95**	5	0.964	-0.008	0.087 - 0.115	0.101	0.058
M5	190.38	24	11.01**	1	0.962	-0.002	0.088 - 0.115	0.102	0.088
M6	215.84	25	25.46**	1	0.956	-0.006	0.094 - 0.120	0.107	0.083

注: * $p < 0.05$, ** $p < 0.01$ 。

从不同组别的完全因素不变性各模型拟合指数来看,均表现出 $CFI > 0.95$ 、 $TLI > 0.95$,大部分 $SRMR < 0.08$,拟合指数较好。具体来看,(1)是否有急事需要处理(不是 vs. 是的),M1 ~ M6 模型差异的 $\Delta\chi^2$ 均表现为 $p > 0.05$,完全因素不变性成立。即,是否有急事需要处理不影响对生活满意度的主观判断。(2)答题时感受(积极情绪 vs. 消极情绪),M1 ~ M4 模型不变性检验的 $\Delta\chi^2$ 均表现为 $p > 0.05$;M5、M6 模型不变性不成立($\Delta\chi^2_{(1)} = 3.96$ 、 20.89 $p < 0.05$)。因此,完全因素不变性部分成立。即,答题时感受积极情绪与答题时感受消极情绪两组在生活满意度的因子方差-协方差不变性与潜均值的不变性不能成立。(3)噪音水平(无噪音 vs. 有噪音),M1、M2 模型不变性检验的 $\Delta\chi^2$ 均表现为 $p > 0.05$;M3 与 M4 不变性检验不成立($\Delta\chi^2_{(4)} = 14.75$ 、 $\Delta\chi^2_{(5)} = 23.91$ $p < 0.05$)。即,无噪音组与有噪音组在生活满意度的截距的不变性与严格不变性不能成立。(4)答题用时(长 vs. 短),M1 ~ M6 模型差异的 $\Delta\chi^2$ 均表现为 $p < 0.05$,完全因素不变性不能成立。也

就是说,答题时间长短会影响对生活满意度的主观判断。

4 讨论

不变检验结果表明,第一,是否有急事需要处理(不是 vs. 是的)的完全因素不变性成立,也就是说这个环境变量不使被试对生活满意度理解产生差异。第二,答题时感受(积极情绪 vs. 消极情绪)、噪音水平(无噪音 vs. 有噪音)的完全因素不变性只有部分成立。其中,答题时感受的因子方差-协方差不变性与潜均值的不变性不成立,这说明积极情绪组潜变量上的离散程度、协方差与消极情绪组不是对等的;噪音水平的截距不变性与严格不变性不成立,这说明无噪音组的截距、误差方差与有噪音组不是对等的。这也说明,答题时感受、噪音水平在网络测验中会对测试结果阐述重要影响,因为会使得被试对生活满意度的理解。第三,答题用时(长 vs. 短)完全因素不变性不成立。这说明,答题用时长组答题用时短组在生活满意度理解完全是不对等的。因此,用时(此次研究使用均值区分)过短的数

据在网络测验中使用要务必小心。实际上,这也说明上述三个指标在网络测验中是对被试答题信效度有效控制的变量。在实际操作中,也不必对于严格不变性、潜变量的均值不变性过分在意,因为相对来说 M1 ~ M3 不变性成立已经满足测量不变性成立的基本要求(Widaman & Reise, 1997)。有研究者认为,多数情况下,对测验只要求具有弱不变性,至多严格不变性(蔡华俭等, 2008)。

关于模型之间比较的检验参数。有研究者认为 ΔCFI 比 $\Delta\chi^2$ 更可靠,因为 $\Delta\chi^2$ 会受到样本量的显著影响,而 ΔCFI 、 ΔTLI 更稳定。当 $\Delta\chi^2$ 与 ΔCFI 不一致时,则建议参考 ΔCFI 结果,当 $|\Delta CFI| > 0.001$ 时,两模型不变性检验假设不成立(Cheung & Rensvold, 2002)。但是, Cheung 和 Rensvold 在研究也明确表示,用 ΔCFI 取代 $\Delta\chi^2$ 是不可取的,也是没有意义。也有研究者认为,当 $|\Delta CFI| > 0.002$ 或者 $\Delta\chi^2 (p < 0.05)$ 时,两模型不变性检验假设就不能成立(Fournier et al., 2007)。因此,如果按照这个标准此次研究的结论也是成立的,故在实际检验中没有把 ΔCFI 作为检验参数,但是研究呈现了 ΔCFI 结果(见表 2)。

不变性检验研究的意义在于,首先,网络测验的情境下如何对测试的质量进行控制,需要增加相关变量进行检测,研究结果表明,测试完成时间、测试时噪音影响、测试时被试情绪等是测试质量的关键变量。对于网络化测试上述问题应该作为可靠性控制必做题。其次,对于完全因素不变性检验做了系统的归纳,将测量不变性、结构不变性与完全因素不变性的概念做了界定与区分,并且系统的归纳了三者的关系,介绍了不变性检验的几种流程,还以 SWLS 为例做了实际的操作验证。研究虽然较为系统的归纳了完全因素不变性检验的方法与步骤,但如果不是专门进行不变性研究报告的撰写应该可以使用更加简洁的方法与步骤,后续研究中可以探索简洁的流程与步骤,实现比较经济且易于学习的效果。另外,对于完全因素不变性的检验参数还有众多不同观点,比如, ΔCFI 作为检验参数仅仅对于两组比较有效,多组比较还需要进一步系统归纳。对于 SWLS 完全因素不变性检验而言,如何实验网络测验的可靠性控制,除了测试完成时间、测试时噪音影响、测试时被试情绪,还有一些因素需要进一步检验,如测试场所(学校、办公室、家里等)、测试网络途径(计算机网络、移动网络)等。这些都是网络测

试带来的新问题,需要在后续研究进一步深入探讨分析。

参考文献

- 蔡华俭,林永佳,伍秋萍,严乐,黄玄凤.(2008).网络测验和纸笔测验的测量不变性研究——以生活满意度量表为例.《心理学报》,40(2),228-239.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255.
- Cieciuch, J., Davidov, E., Oberski, D. L., & Algesheimer, R. (2015). Testing for measurement invariance by detecting local misspecification and an illustration across online and paper-and-pencil samples. *European Political Science*, 14(4), 521-538.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16-29.
- Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods Instruments & Computers: A Journal of the Psychonomic Society Inc*, 31(4), 572-577.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71-75.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125(2), 276-302.
- Dodou, D., & Winter, J. C. F. D. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, 36(36), 487-495.
- Fournier, J., Gaudreau, P., Demontrond-Behr, P., Visioli, J., Forest, J., & Jackson, S. (2007). French translation of the Flow State Scale-2: Factor structure, cross-cultural invariance, and associations with goal attainment. *Psychology of Sport and Exercise*, 8(6), 897-916.
- Hardre, P. L., Crowson, H. M., & Xie, K. (2012). Examining Contexts-of-Use for Web-Based and Paper-Based Questionnaires. *Educational & Psychological Measurement*, 72(6), 1015-1038.
- Hauk, S. (2015). A Comparison of Web-based and Paper-

- and - Pencil Homework on Student Performance in College Algebra. *Primus Problems Resources & Issues in Mathematics Undergraduate Studies* 25(1) 61 - 79.
- Lewis ,M. I. ,Watson ,B. & White ,K. M. (2009). Internet versus paper - and - pencil survey methods in psychological experiments: Equivalence testing of participant responses to health - related messages. *Australian Journal of Psychology* , 61(2) ,107 - 116.
- Meade ,A. W. & Lautenschlager ,G. J. (2004). A Monte - Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance. *Structural Equation Modeling A Multidisciplinary Journal* ,11(1) 60 - 72.
- Meade ,A. W. ,Lautenschlager ,G. J. & Michels ,L. C. (2007). Are Internet and Paper - and - Pencil Personality Tests Truly Comparable? An Experimental Design Measurement Invariance Study. *Organizational Research Methods* ,10(10) 322 - 345.
- Raffaelli ,M. ,Armstrong ,J. ,Tran ,S. P. ,Griffith ,A. N. ,Walker ,K. ,& Gutierrez ,V. (2016). Focus on Methodology: Beyond paper and pencil: Conducting computer - assisted data collection with adolescents in group settings. *Journal of Adolescence* #9 ,1 - 9.
- Vandenberg ,R. J. ,& Lance ,C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions , practices ,and recommendations for organizational research. *Organizational Research Methods* 3(1) 4 - 70.
- Wang ,K. T. ,Yuen ,M. & Slaney ,R. B. (2009). Perfectionism , depression ,loneliness ,and life satisfaction a study of high school students in Hong Kong. *The Counseling Psychologist* , 37(2) 249 - 274.
- Widaman ,K. F. & Reise ,S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research* , (9) 281 - 324.

Testing Measurement Invariance of Satisfaction with Life Scale (SWLS) based on the Multi - group Confirmatory Factor Analysis

Wang Daoyang^{1,2} ,Liu Zhengguang²

(1. Department of Psychology ,Anhui Normal University ,Wuhu 241000;

2. Collaborative Innovation Center of Assessment toward Basic Education Quality ,Beijing Normal University ,Beijing 100875)

Abstract: Measurement Invariance is important in the psychometric application of self - report scale or questionnaire ,which is a prerequisite for cross - group comparison. As an popular method to analyze Measurement Invariance ,multi - group confirmatory factor analysis contains various theoretical models ,and has different combination of models in application. The current study systemetically summarized these relative theoretical models and then proposed several models for testing Factorial Invariance ,including confirmatory factor analysis across groups without any limitation (Mgroup) ; Configural Invariance Model (M1) ; Metric Invariance Model (M2) ; Scalar Invariance Model (M3) ; Strict Invariance Model (M4) ; Factor Variance / Covariance Invariance Model (M5) ; Latent Mean Invariance Model (M6). The study took the Satisfaction With Life Scale (SWLS) as an example ,with the subjects of 1343 college students (age: 17 - 25 , 20.01 ± 1.53) ,to test SWLS Measurement Invariance according to different groups as below: In a hurry or not (Hurry vs. No - hurry) , Mood in the test (Positive mood vs. Negative mood) ,Noise or not (Noise vs. Non - noise) ,Response time (Long response time vs. Short response time) ,Gender (Male vs. Female) ,Hukou (Agricultural vs. Non - agricultural). The results show that: a) Factorial Invariance is tenable across the Hurry group and No - hurry group ($\Delta\chi^2 = 0.49 \sim 10.59$, $p > 0.05$) ; b) Partial Factorial Invariance is tenable across the Positive mood group and Negative mood group ,with M5 Invariance Model and M6 Invariance Model untenable ($\Delta\chi^2_{(1)} = 3.96$, 20.89 $p < 0.05$) ; c) Partial Factorial Invariance is tenable across the Noise group and Non - noise group ,with M3 Invariance Model and M4 Invariance Model untenable ($\Delta\chi^2_{(4)} = 14.75$, $\Delta\chi^2_{(5)} = 23.91$, $p < 0.05$) ; d) Factorial Invariance is tenable across the Long response time group and Short response time group ($\Delta\chi^2 = 11.01 \sim 41.95$, $p < 0.05$) ; e) Partial Factorial Invariance is tenable across the Male group and Female group ,with M4 Invariance Model untenable ($\Delta\chi^2_{(5)} = 64.40$, $p < 0.05$) ; f) Partial Factorial Invariance is tenable across the Rural group and Non - rural group ,with M6 Invariance Model untenable ($\Delta\chi^2_{(1)} = 11.49$, $p < 0.05$).

Key words: measurement equivalence; structural invariance; SWLS